14th ICCRTS:
"C2 and Agility"

Event Detection Challenges, Methods, and Applications in Natural and Artificial Systems

Topic Categorization:
Modeling & Simulation

Mitchell C. Kerman
System of Systems Engineering
Lockheed Martin MS2
199 Borton Landing Road
Moorestown, New Jersey 08057
mitchell.c.kerman@lmco.com
(609) 326-5156

Wei Jiang, Ph.D.
School of Systems & Enterprises
Stevens Institute of Technology
Castle Point on Hudson
Hoboken, New Jersey 07030

Alan F. Blumberg, Ph.D.
Director, Center for Maritime Systems
Stevens Institute of Technology
Castle Point on Hudson
Hoboken, New Jersey 07030

Samuel E. Buttrey, Ph.D.
Operations Research Department
Naval Postgraduate School
1 University Circle
Monterey, California 93943

March 2009

© 2009 Lockheed Martin MS2

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Event Detection Challenges, Methods, and Applications in Natural and Artificial Systems	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,Operations Research Department,1 University Circle,Monterey,CA,93943**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
In Proceedings of the 14th International Command and Control Research and Technology Symposium (ICCRTS) was held Jun 15-17, 2009, in Washington, DC.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**58**	

Event Detection Challenges, Methods, and Applications in Natural and Artificial Systems

Mitchell C. Kerman, Lockheed Martin MS2
Wei Jiang, Stevens Institute of Technology
Alan F. Blumberg, Stevens Institute of Technology
Samuel E. Buttrey, Naval Postgraduate School

Abstract

. . . .

A system is a combination of elements whose collaborative actions produce results generally not attainable by the elements acting alone, and an event is a significant occurrence or large-scale activity that is unusual relative to normal patterns of behavior. Event detection, or the process of identifying the occurrence of an event, within both natural and artificial (or man-made) systems has long been a topic of research, and a variety of techniques have been developed to address event detection problems.

This article is a treatise on the topic of event detection and a prequel to research previously conducted by the authors regarding the application of robust metamodels to uncertainty quantification and event detection within a geophysical system. The article explores the most common difficulties and challenges in event detection problems, describes the event detection methods most frequently employed, and provides example event detection applications in both natural and artificial systems. It incorporates the discoveries of and lessons learned by multiple researchers and authors over many combined years of experience in event detection theory and application; this rather broad study has never been previously published within a single volume. The article concludes with an examination of the intimate relationship and indivisible link between event detection and modeling and simulation.

. . . .

© 2009 Lockheed Martin MS2

Introduction

As defined by the International Council on Systems Engineering (INCOSE), a system is "a combination of interacting elements organized to achieve one or more stated purposes" [INCOSE, 2006]. Expanding on this definition, a system can be described as a collection of elements that, in combination, produce results generally not obtainable by the elements acting alone. These elements may include operators, hardware, software, firmware, information, policies, documents, techniques, facilities, services, and other support components; that is, all items required to produce system-level results. System-level results are the qualities, properties, characteristics, functions, behaviors, and performance of the entire system. Thus, it is the interconnection and interaction of the individual system elements that delineate the system-level results, producing a desired behavior beyond the capacity of any individual system element or subgroup of system elements.

There are numerous types of systems and a myriad of ways in which to categorize the types of systems. In the most basic classification scheme, systems may be simply categorized into natural and artificial (or man-made) systems. While natural systems may not have an apparent objective, the system inputs and outputs can be interpreted as serving a purpose. Artificial systems, on the other hand, are designed for a specific purpose which is achieved through the delivery of outputs or services. Systems may also be subcategorized into observable and non-observable systems. An observable system is one in which the system inputs and outputs may be directly perceived in real-time. In non-observable systems, however, either or both the system inputs and outputs may not be directly observed. Finally, all systems can be analyzed both qualitatively and quantitatively – qualitatively through the delivery of the outputs and quantitatively through the measurement and analysis of specific system performance and effectiveness metrics derived from the system outputs.

An event is a significant occurrence or large-scale activity that is unusual relative to normal patterns of behavior. Examples of events include a large meeting being conducted in an office building, a malicious attack on a Web server, or a traffic accident occurring on a freeway [Ihler, Hutchins, and Smyth, 2006]. In terms of systems, events may be associated with naturally occurring phenomena and manual system interactions. Some naturally occurring phenomena include chemical and thermodynamic reactions and physical processes in the time and space domains. An operator pushing a button is an example of a manual system interaction resulting in a "button pressed" event. Generally, an event results in the aberration of system parameters and output metrics. Therefore, events may be identified through a process known as event detection.

For parameters of observable systems, event detection is usually a simple matter of observing the system states. For example, one can easily detect whether or not it is raining (in his current locale) by examining the weather outdoors. For non-observable systems or system parameters that are not directly observable, sensors are normally employed to track the states of the parameters of interest. While one can easily determine the relative temperature of his surroundings, for instance, he must employ thermometers and temperature gauges (i.e., temperature sensors) to ascertain whether or not the temperature is below the freezing point of water.

Sensors may be organic (to the detection platform), local, remote, or any combination thereof, and the sensor outputs are used as the inputs to the event detection systems. In both

natural and artificial systems, however, sensor-based event detection is among the most difficult and time-constrained of analysis problems, typically requiring excessive computational power and large amounts of storage space for voluminous data. Examples of significant events which are normally detected using sensor-based event detection include a substantial change in sea level, an increase in background radiation level, the maneuver (or course change) of an anti-ship missile, and an increase in pressure within a boiler (or heat exchanger).

Various methods of sensor-based event detection have been developed, and static threshold event detection is one of the simplest and most commonly used of these methods. For example, a fuel level sensor in an automobile's gasoline tank can easily detect a low fuel condition when the fuel level falls below a set (i.e., static) threshold value. The detection of this condition (or event) activates a visual indication (e.g., a warning light) and (in many vehicles) an audible alarm to alert the operator of the low fuel condition. Since static threshold event detection is a relatively simple method, it is typically less reliable than more advanced techniques. If the fuel level sensor fails, for instance, the operator may receive a false low fuel level warning or, even worse, no warning indication of an actual low fuel condition. To overcome the reliability issues associated with a single sensor, many systems employ multiple (or redundant) sensors. Additionally, systems that must detect events over large swaths of time and space typically employ multiple sensors distributed throughout these domains. Of course, multiple sensors add complexity to the event detection problem since multiple inputs must be evaluated in order to determine whether or not an event is transpiring.

Geophysics is the physics of the earth and its environment, including the physics of fields such as meteorology, oceanography, and seismology. Thus, geophysical systems are earth-centric systems, such as weather systems, coastal and ocean water systems, and ecological systems. Prior articles and research specifically address uncertainty quantification (UQ) and event detection within oceanographic (or geophysical) data. Kerman, Jiang, Blumberg, and Buttrey [2008] demonstrate an innovative technique for the UQ of salinity data supplied by the New York Harbor Observing and Prediction System (NYHOPS). A continuation of this research utilizes a robust UQ metamodel for event detection within NYHOPS salinity data. Event detection results from a static threshold method are compared against those of a dynamic UQ-based technique and a composite technique that combines both the static and dynamic methods. The results clearly show a significant reduction in the number of false positive detections when using the composite event detection method [Kerman, Jiang, Blumberg, and Buttrey, 2009].

Although the techniques and utility of the aforementioned method have been clearly demonstrated, there is still much work and research to be conducted within the realm of event detection. This article provides an overview of event detection challenges, methods, and applications in natural and artificial systems. While a definitive reference on the topic of event detection would surely comprise multiple volumes, the goals of this article are to introduce the reader to the most common difficulties and challenges in event detection problems, describe the event detection methods most frequently employed, and provide example event detection applications.

The next section describes common challenges in event detection. The third section categorizes and summarizes typical event detection methods, and the fourth section examines example event detection applications. The last section describes the intimate relationship and

indivisible link between event detection and modeling and simulation. Finally, the paper concludes with remarks regarding the future of event detection and its related research.

Common Challenges in Event Detection

The complexities of event detection problems pose an array of challenges. Regardless of the specific detection problem and field of study, there are several common challenges and universal truths in the development and application of event detection methods. The paragraphs that follow summarize the discoveries of and lessons learned by multiple researchers and authors over many combined years of experience in event detection.

Situational Dependence

Event detection problems are extremely situationally-dependent. Several problems may be similar, but no two problems are ever exactly the same. The parameters, variables, and output metrics are selected based upon the specific event detection problem under examination, and these artifacts may or may not be applicable to other problems within the same domain, even for very closely-related problems. However, the approach to address an event detection problem in one domain may inspire alternative or never-before-attempted methods within other domains.

Criticality of Application

Event detection problems are often based upon the requirements of a critical application. Examples of critical applications include monitoring critical assets, measuring indicators of imminent catastrophic machine failures, detecting breaches within security perimeters, and observing human stasis parameters (or vital signs). Such applications typically require a high degree of precision and extreme timelines. High precision means that the event detection method provides a high true positive (i.e., correct detection) rate while also providing a low false positive (i.e., incorrect detection) rate [Dash, Margineantu, and Wong, 2007]. An extreme timeline is a timeline in which the event detection method must be able to correctly identify events within a very short amount of time. Depending upon the application, an extreme timeline may range from less than a second to several minutes in duration. Thus, the event detection method must operate in real-time and fast enough to address the criticality of the application so that the detection report is not too time-late for an action or reaction to occur.

Numerous and Diverse Data Sources

The digital revolution has effectively exploded the number of data sources and the amount of data readily available. Unstructured data, text documents, images, audio, video, relational data, multivariate records, and spatio-temporal data are now easily created, stored, and transmitted via computers and other digital electronic devices, such as digital cameras and cellular phones. Any single event detection problem may consider a variety of these diverse data sources consisting of different data types and formats. Thus, the event detection problem is compounded by the predicament over determining what data is actually relevant to the event detection problem under study, and the event detection approach must be capable of evaluating the data from the selected sources.

© 2009 Lockheed Martin MS2

Event detection problems typically involve enormous volumes of data, often measuring in terabytes. Even after reducing the amount of data by selecting specific data sources, the data sets for event detection problems can still be exceptionally large. High-powered computing machinery and immense digital storage space is typically required in order to store, access, filter, and process all of the data within a reasonable timeframe. Even with today's high-speed computers, however, larger data sets still require lengthy processing times.

The overall size of the data set is also of concern for analysis reasons. While vast data sets entail large storage and processing overheads, too little data is also problematic. Too little data can lead to missed detections or the development of an event detection solution which does not work in all cases. Too much data, on the other hand, can lead to "analysis paralysis," a situation in which the detection problem is (over-)analyzed over the course of many years but never really solved using a practical event detection approach due to the overwhelming volume of data.

The data itself may be provided by a single sensor, an array of identical sensors, or an inhomogeneous mix of sensors. Consequently, data gathered to address the same event detection problem may originate from a variety of sensors using different information storage, retrieval, and transmission formats. The event detection algorithm must transform the low-level sensor data into high-level events. In order to accomplish this, the algorithm must aggregate, convert, or reformat the received data into a uniform structure that is independent of the data source [Fienberg and Shmueli, 2005].

Raw sensor data is often plagued by inaccuracies and incompleteness. For instance, position information may be inaccurate or missing. Furthermore, information transmitted from mobile devices can be delayed and arrive at the detection system out-of-order. Environmental information, for example, may be approximated from out-of-date reports of mobile sensors in the area of interest [Balazinska, 2007]. Raw sensor information may have confidence bounds, and environmental activity information may also have limited confidence. Thus, the detection system must consider the underlying inaccuracies, incompleteness, and confidence levels of the raw data when determining the presence of events.

Raw data from environmental sources often exhibits cyclical, seasonal, and irregular trends. These trends must be understood and considered within the event detection framework. In addition, raw data is often corrupted by a number of "burst" periods of atypical or unusual behavior as in the example cases of the number of customers entering a bank or the number of freeway traffic accidents. The raw data consists of the aggregated behavior of the individual elements of the system through the interdependencies and interactions of these elements. Therefore, any truth data gathered from the actual system reflects the rhythms of these underlying activities, and the resulting data set will appear non-homogeneous. This leads to an inherent "chicken and egg" deconvolution problem. The presence of large events distorts the estimated rate of "normal" behavior, resulting in a slight increase in the detection parameters which causes the static event detection thresholds to miss the presence of other events (around the same time or in the same locale). "Detecting anomalous periods of time requires some knowledge of what constitutes normal behavior, but historical data consists of both normal and anomalous (event) data mixed together" [Ihler et al., 2006]. The event detection system must examine the patterns of typical and predictable behavior as well as detecting and extracting information from the deviations from this behavior. Ideally, the system should "learn" the

patterns of normal behavior and have a method to detect events that indicate departures from the norm.

Network Topology

A network is a system of interrelated stations spread throughout a region or area. In terms of this study, a network is a system containing a number of transmitting and receiving sensor stations, or nodes, that are connected through cables, wires, or wireless communications medium. Network topology considers the locations and connectivity of these sensors in relation to the entire sensor network over time. In remote and mobile sensor networks, the network topology changes continuously due to sensor mobility and sensor lifetime. Remote sensor nodes within a wireless sensor network, referred to as "motes," require maintenance and reseeding due to movement outside of the intended observed area, power consumption, sensor failures, and finite sensor lifetimes. So, the care and maintenance of the sensor network itself constitutes another challenge in event detection.

The communications medium utilized in mobile and remote sensor networks may also have limits on throughput and capacity. Thus, the network design engineers face a critical decision as to what type of data to transmit across the sensor network, full sensor-level data or spatially-temporally aggregated data. "Aggregated data are simpler to transmit and raise limited privacy [and security] issues. Individual level data are more voluminous, raise numerous privacy [and security] concerns, and may not easily be merged because of differences in forms of identifiers in diverse systems as well as errors" [Fienberg and Shmueli, 2005]. Obviously, the most informative sensor data are those that are collected within the influence domain of the event. Aggregating data may increase network throughput and reduce data processing times, but it can significantly reduce the chance of detection since data from unaffected areas can mask the event signature. Thus, it will take longer for the detection system to notice the slight change in the aggregated data. Collecting localized, sensor-level data may be more beneficial for improving detection sensitivity, but the processing times for extracting and evaluating the larger volumes of data can also affect the timeliness of detection.

The delay in receipt of raw sensor data is briefly discussed in the previous section. It is usually not feasible to defer composite event detection and wait for all delayed sensor reports from remote sites, but delayed sensor reports can affect *event persistence* and *event lifetime*. Event persistence is, essentially, the number of positive sensor detections required (from the same sensor) in order to report the occurrence of an event. Event lifetime is the length of an event as determined by the event persistence algorithm in signaling the start and end of the event. Network delays in the receipt of sensor reports can also result in event "context fluttering." This is the situation in which the event context (i.e., the indication of an event) is activated and deactivated in close succession due to inaccurate sensor readings or network delays. Within a single sensor, activation and deactivation in close succession due to errors is known as sensor "hunting." The challenges of event context fluttering consist of detection and prevention measures. The event detection algorithm must be able to detect the presence of event context fluttering and then deal with the situation. Prior solutions have incorporated hysteresis parameters in the computation of the event detection threshold setpoints [Schwiderski-Grosche, 2008].

© 2009 Lockheed Martin MS2

Event Detection Algorithms

The event detection approach and resulting algorithm must address and overcome the aformentioned challenges as well as several others. The three main requirements of an event detection algorithm are timeliness, a high true detection rate, and a low false alarm rate.

Many event detection algorithms require some time for the algorithm parameters to properly initialize, learn from the event-free environment, and then reach a stable state. Thus, algorithm time considerations for critical event reporting must factor in the algorithm initialization, learning, and stabilization times. Preliminary (learning) data for the algorithm must be known to be void of the events of interest. Thus, the detection system "learns" to detect the event based upon the event-free situation. The disadvantage of this learning process, however, is that some events may not be immediately recognized by the algorithm. A "day zero" event, for instance, is an event which is uncharacteristic of the normal events and has a never-seen-before signature. The detection algorithm has no means to detect a "day zero" event, as the algorithm is not actively "looking" for it. Operator interaction or intervention is usually required in order to recognize and report the event. Based upon the experience of Fienberg and Shmueli [2005], it is not possible to achieve a fully automated event detection system. Careful selection of methods and algorithms can minimize the need for operator intervention at the cost of a reduction in discrimination power.

The timeliness requirement for the event detection algorithm also implies immediate analysis of incoming data and immediate reporting of the results. To analyze data quickly, the incoming data must be stored in a format that the detection algorithm can use and the algorithm must be computationally efficient to run quickly. The processing time for extracting data in a useful format and carrying out even the simplest calculations can be prohibitively long. While there is value in using high sample frequency sensor-level data, the development of efficient methods for handling this data is still a challenge. Fast storage and analysis are critical when dealing with massive data sets. So, it is essential that the detection algorithm be efficient (i.e., fast and computationally cheap). Thus, the timeliness requirement may give priority to some solution approaches over others.

The detection algorithm should integrate all previous data and use it to decide whether or not a new data point indicates the presence of an event. This approach is known as "roll-forward." Each new data point is assessed for the indication of an event as it is added to the data set. Upon completion of the analysis, the detection system should output an operational decision-making conclusion. "Since the users of the system will usually not be statisticians, the output must be in a user-friendly format which can be easily understood [and transferred, such as in graphs, charts, or reports]. Of course this output should be immediate and not delay the process of decision making" [Fienberg and Shmueli, 2005].

In the presence of uncertainty in the input data, the detection algorithm must make a trade-off between event *precision* and *recall*. Precision is the fraction of reported events that are actual (true) events. Recall is the fraction of all events that are reported correctly. In a pessimistic approach, the algorithm ignores large numbers of (potential) events due to the data uncertainty. While the precision (or correct reporting) of the algorithm is high using this approach, many actual events are missed, reducing the recall value. In the opposite case, the optimistic approach reports events even in the presence of the uncertain data. Here, the precision of the algorithm is lower since the errors in the input data result in false events, but the recall

value is higher since fewer true events are missed [Balazinska, 2007]. "Current systems tend to have a high true-detection rate, but at the cost of a high false-alarm rate" [Fienberg and Shmueli, 2005]. However, "all systems can be calibrated to have less false alarms at the expense of a slower true detection rate, and vice-versa. Although the risk of not detecting a true [event] should be minimized, the cost and handling of false alarms must be taken into account" [Fienberg and Shmueli, 2005].

Finally, many event detection problems are exacerbated by the presence of an active adversary. This adds a level of complexity to the event detection algorithm. For example, an enemy submarine will covertly hide from a surface ship that is attempting to detect it.

Typical Event Detection Methods

While there is no clear manner in which to characterize every event detection method, typical event detection methods may be classified into four rather broad categories: statistical, probabilistic, artificial intelligence and machine learning, and composite. Most event detection methods fit into one of these categories, but several methods encompass two or more. This section describes several event detection methods and provides the associated references. By no means is this an exhaustive reference; this section just provides example methods within each category. As the theory and implementation details of most of these methods are rather lengthy, the reader is encouraged to consult the cited references for this information.

Statistical Methods

Static threshold event detection is by far the simplest and most computationally straight-forward method of statistical event detection. Event detections are reported when the monitored parameter exceeds a predetermined threshold value, and the detection condition persists as long as the parameter value exceeds the threshold setpoint. Once the parameter falls within acceptable bounds (e.g., below the threshold value), the detection condition clears. Threshold values may be determined based upon historical parameter values, analogy to similar sensors and systems, engineering estimates, or parametric analysis. The static threshold method exhibits a "memoryless" property from one observation to the next, as the current observation and detection condition is independent of all prior observations. However, observed values are (usually) dependent upon prior observed values, and one would not reasonably expect the observed values to radically change in the short period of time between successive observations. Many references describe the benefits and utility of static threshold event detection methods, and Kerman et al. [2009] baseline the results of the static threshold method against a composite event detection method.

Regression is a data modeling and analysis technique in which the dependent variable is modeled as a function of independent variables, constant parameters, and an error term. The error term represents the variation in the dependent variable that cannot be explained by the model, and it is modeled as a random variable [Sykes, 1993]. Regression is often used for forecasting and prediction, inference, hypothesis testing, and exploring relationships among data parameters [Wikipedia, 2008].

Linear regression models the relationship between the dependent and independent variables as a straight line. When the relationship between the dependent and independent variables is clearly non-linear, polynomial regression may be used to provide a better-fitting

polynomial model. Sauvageon, Agogino, Mehr, and Tumer [2006], for instance, use a fourth degree polynomial within an event detection algorithm to sense high temperatures on an aluminum plate. LOESS regression is another one of the various regression techniques, and it can be described as locally weighted scatterplot smoothing. This technique combines the simplicity of linear least squares regression with the flexibility of nonlinear regression. Simple models are fit to local subsets of data in order to create the LOESS regression function that describes the deterministic part of the variation in the data, point by point. "In fact, one of the chief attractions of this method is that the data analyst is not required to specify a global function of any form to fit a model to the data, only to fit segments of the data" [NIST, 2008]. Quantile regression (QR) is another regression method, and it estimates models for the full range of conditional quantile functions, including the conditional median function, thereby providing a more complete statistical analysis of the stochastic relationships among the random variables.

Time series analysis consists of statistical modeling methods for data that is typically measured at successive times in equally spaced time intervals. These models may have many different forms to represent the various underlying stochastic processes. The three general types of time series models include autoregressive (AR), integrated (I), and moving average (MA). Time series techniques may also be used in combination to yield such models as autoregressive-moving average (ARMA) and autoregressive-integrated-moving average (ARIMA).

Kalman filtering is another statistical modeling technique. "The Kalman filter is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown" [Welch and Bishop, 2006]. Kalman filters are commonly used in a wide range of engineering applications, including radar tracking and computer vision algorithms. The Kalman filter also forms an important topic in control theory and control systems engineering.

Kerman et al. [2008] explore the UQ of salinity metamodels developed using a variety of statistical techniques, including LOESS regression and quantile regression, autoregressive and moving average time series models, and Kalman filtering. The uncertainty of these metamodels is investigated for their expected accuracy and applicability within a real-time salinity event detection system. The authors continue this research by describing a quantile regression metamodel for use within a salinity event detection algorithm [Kerman et al., 2009].

Sauvageon et al. [2006] use model fitting interpolation in a sensor network analysis for detecting surface temperature change on an aluminum plate. The interpolation method presented is the bicubic technique, the most common interpolation method in two dimensions. This technique consists of two basic cubic interpolations, one in each plane direction. At each point, the value of the function is computed as the weighted average of its nearest sixteen neighbor points.

Lastly, Gupchup, Burns, Terzis, and Szalay [2007] apply statistical signal processing techniques to event detection in wireless sensor networks. Specifically, they use Principal Component Analysis (PCA) to build a model of observed environmental phenomena that captures daily and seasonal trends within the sensor measurements. The divergence between sensor measurements and model predictions is used as an indicator of discrete events within the data stream. PCA, also known as the Karhunen-Loève transform (KLT), "is a powerful

statistical tool for simplifying data by reducing high-dimensional datasets into datasets with lower dimensions that approximate the original data. It does so through singular value decomposition (SVD): an orthogonal linear transform of a matrix containing the original data into an equivalent diagonalized matrix" [Gupchup et al., 2007].

Probabilistic Methods

Probabilistic event detection methods consist of those methods in which the probability of event occurrence and other related probabilities and parameters are computed and assessed rather than computing and testing statistics from a sample data set. Ihler et al. [2006], for instance, develop a probabilistic framework for unsupervised event detection and learning based upon a time-varying Poisson process model that can also account for anomalous events. Their experimental results indicate that the proposed time-varying Poisson model provides a robust and accurate framework to adaptively separate unusual event plumes from normal activity. This model also performs significantly better than a non-probabilistic, threshold-based event detection technique.

Sauvageon et al. [2006] investigate the Distributed Gaussian Method (DGM) for detecting surface temperature changes. In this technique, Gaussian curves are generated such that they are centered on each node. Then, these curves are normalized and summed in order to reduce the geometric effect of node placement. The maximum predicted temperature value is then easily located in order to detect the temperature peak.

Grid computing involves groups of heterogeneous computational servers connected via high-speed network connections. Tham [2006] describes SensorGrid, an architecture for integrating sensor networks with grid computing. In this architecture, real-time information about phenomena in the physical world can be mined, extracted, correlated, and processed to facilitate "on-the-fly" decisions and actions in response to real-world events. This architecture relies upon distributed data fusion, event detection, and classification via probabilistic algorithms.

Artificial Intelligence and Machine Learning Methods

Artificial intelligence (AI) and machine learning (ML) event detection methods are usually both computationally and informationally intensive. The sensor sources in the modeled systems are often sparsely distributed in time and space. Thus, these methods require advanced fusion algorithms in order to correlate the data from multiple sources.

Database operations, such as queries and table joins, are among the most direct of these methods. Abadi, Madden, and Lindner [2005] develop REED, a system for robust, efficient filtering and event detection in sensor networks. REED consists of a set of algorithms that efficiently evaluate join queries on static data tables in a sensor network. Their solution allows complex time and location queries by storing the filter conditions in tables and then distributing the tables throughout the network.

The Mote Fuzzy Validation and Fusion (Mote-FVF) algorithm was developed for wireless sensors network. This algorithm can distinguish between sensor failures and abnormal environmental behaviors by using network redundancy to compensate for sensor reliability. Fuzzy logic based methods for sensor validation and fusion are unique in that they do not require or rely upon a mathematical model of the system. Among other algorithms, Sauvageon et al.

© 2009 Lockheed Martin MS2

[2006] also test the Mote-FVF algorithm for its performance in detecting high temperatures on an aluminum plate.

There are various other AI and ML methods that are directly applicable to event detection but are not directly addressed in this exposition; such methods include particle filtering, genetic algorithms, neural networks, and intelligent agents. A simple Internet search for each of these methods in regard to event detection yields numerous results and multiple references.

Composite Methods

Composite event detection methods are those methods that combine techniques within a category or from two or more of the categories. Bayesian Gaussian Process (BGP) models, for instance, combine probabilistic and machine learning methods. BGP classification techniques are powerful non-parametric learning methods based on simple probabilistic models. BGP modeling is a stochastic process which generates samples over time. Regardless of the finite linear combination of random variates selected, the resulting linear combination will be normally distributed. Kerman et al. [2008] explore a BGP metamodel for the UQ of NYHOPS salinity data.

Kerman et al. [2009] also develop a composite event detection technique for salinity event detection confirmation within New York Harbor oceanographic data. Their method combines the static threshold method with a dynamic UQ event detection technique based upon quantile regression. The results of the composite event detection method show a significant reduction in the number of false positive detections.

Example Event Detection Applications

In general, in order to detect an event, one must be actively looking for it. In other words, the correct data must be collected, examined, and analyzed in order for an event to be detected. Given that it is still difficult to detect an event even while actively seeking it, the chances are slim that an event will be detected by happenstance alone. Therefore, it makes sense that the primary purposes of event detection are monitoring, surveillance, and management of systems and processes. This section examines example event detection applications and classifies and organizes these applications according to their problem domains. Sample reference documents are also cited when applicable.

Network Monitoring

In today's information-intensive world, network monitoring is of paramount importance. For instance, businesses are often interested in the frequency of visits to their websites and the general geographic locations of the visitors. Additionally, businesses are concerned with Internet usage by their employees. Monitoring Internet connections and conducting Web access logging is necessitated by both of these applications. Furthermore, banks are interested in maintaining the security of their online systems for their protection and the protection of their customers. So, website intrusion detection and failed account access logging are standard security practices [Gupchup et al., 2007].

Traffic monitoring is another network monitoring application. For instance, a community may be interested in monitoring traffic at an intersection in order to determine whether or not it

warrants a new traffic signal. Ihler et al. [2006] use freeway traffic data as a test case in their development of an adaptive event detection method based upon a time-varying Poisson process.

Health Monitoring and Management

The detection and prediction of conditions or events is also of extreme importance in healthcare applications. The Center for Disease Control and Prevention (CDC), for instance, continuously monitors medical and public health information from physicians and hospitals across the country. This practice is necessary for the earliest possible detection of viruses and disease. Any detected wide-spread illness must be contained by quarantining and treating the afflicted individuals. The objective is to prevent further spreading of the illness so that it does not result in an epidemic or, worse, a pandemic. Afflictions of interest are usually naturally occurring such as the influenza virus, but bio-terrorist attacks present another area of grave concern [Fienberg and Shmueli, 2005].

The early detection of disease within individual patients presents another health management issue. Zelen [2007] discusses screening and monitoring programs for the early detection of diseases such as diabetes, hypertension, thyroid disease, tuberculosis, cancer, and coronary artery disease. His study examines the age at which to begin screening exams, the intervals between the exams, and (possibly) the age at which to end screening exams. The motivation of the study is the early diagnosis and treatment of disease before the patient shows any signs or symptoms (i.e., when the patient is in the pre-clinical state). The hope is that early treatment will result in more cures and a lower mortality rate. In a similar study, Favretto, Farias, and Murta [2007] examine the main aspects of a decision support system for automatic detection of ischemic events within digital electrocardiogram (ECG) data. Ischemia is a condition defined by insufficient blood flow to maintain tissue's normal function, and myocardial ischemia is one of the most significant heart diseases.

Health monitoring and management is also of critical importance in aerospace applications. “Event detection using centralized sensor networks is often regarded as one of the most promising health management technologies in aerospace applications where timely detection of local anomalies has a great impact on the safety of the mission” [Sauvageon et al., 2006].

Environmental Monitoring and Prediction

Environmental monitoring and prediction is another common area for the application of event detection methods. The earth's environment can be extremely violent, and early warnings of impending natural disasters such as tornadoes, hurricanes, tsunamis, earthquakes, floods, and volcanic eruptions are critical for the safety and security of populations within the affected regions. For example, Hurricane Ike recently devastated the city of Galveston, Texas. Due to the influence of early detection and warning systems, the majority of the populace was safely evacuated prior to hurricane landfall [MSNBC News Service, 2008]. Additionally, the contamination of natural resources, whether it be by natural or man-made (e.g., terrorist) causes, is another area of concern. Potable water, for instance, is continuously monitored by water utilities for purity and potential contaminants.

One example of environmental monitoring is a study regarding the detection of salinity events within oceanographic data [Kerman et al., 2009]. As a second example, Trafalis, Ince,

and Richman [2003] devise a method to more accurately detect tornadoes by using Support Vector Machines (SVMs). In comparison with other detection methods, such as neural networks and radial basis function networks, SVMs are found to be more effective in tornado detection.

Safety and Security

Safety and security applications are other areas that utilize event detection methods. For instance, physical intrusion detection and fire safety are of critical importance to businesses and homeowners. Automobile, home, and corporate security alarm systems deter potential thefts and mischievous acts. Furthermore, fire, smoke, and carbon monoxide alarm systems increase survivability in the advent of a fire or buildup of toxic gas.

Developing a prediction method for 9-1-1 call volumes can aid emergency service providers in service planning and recognition of anomalous calls. This topic was researched by Jasso, Fountain, Baru, Hodgkiss, Reich, and Warner [2007] in analyzing emergency events in the San Francisco Bay area.

Homeland Security plays an important role in today's security environment, and cargo security is a main area of concern. Shippers and Customs and Border Protection agents are concerned with verifying that the contents of cargo was not compromised during shipment. Furthermore, the military is concerned with threat detection and management. The detection, tracking, and interception of threat missiles is a quintessential military threat management example.

Tavakoli, Zhang, and Son [2005] examine a group-based event detection architecture for undersea wireless sensor networks that communicate via underwater acoustics. This architecture can be used to detect underwater intrusions such as those posed by enemy submarines within an operating area. The research indicates that false alarms are one of the main problems in this type of surveillance application, and the authors develop a group-based event detection algorithm to reduce the number of false alarms. Spatial and historical information is used to increase the confidence level of the event reports by assigning larger confidence values when a higher percentage of sensors detect the target or when sensors detect the target for a number of consecutive time slots.

Business Process Optimization

Manufacturers rely heavily upon event detection methods to reduce overall maintenance costs and ensure compliance with requirements. Manufacturing and condition-based maintenance is one example. In industrial plants, engineers are concerned with identifying machines or processes that are in need of repair or adjustment. Business process compliance is another issue. In food and drug manufacturing, strict regulatory requirements obligate companies to certify that their products do not exceed specific environmental parameters during processing [Abadi et al., 2005].

Furthermore, today's fast-paced and constantly evolving high-tech business environment is extremely demanding on organizations and requires a detailed visibility into real-time business activities. Thus, the ability to efficiently detect correlated business process events that represent opportunities or problems to the organization and require quick action from the decision-maker is paramount.

© 2009 Lockheed Martin MS2

Event Detection Modeling and Simulation

Certainly, there is a strong affinity between event detection and modeling and simulation (M&S). In fact, there is an intimate relationship and indivisible link between the two. Initially, M&S may be used at the forefront of the systems engineering process as a requirements development tool for an event detection system. The M&S development requires a detailed study of the real-world system, thereby facilitating deeper insight into the system interactions, examination of the parameters of interest, and understanding of the relationships between the system inputs and outputs. Thus, it is through M&S that the systems engineer may determine what events can and need to be detected and what parameters must be monitored in order to detect these events. For example, an engineer may determine that mechanical vibration and noise levels must be monitored as indications of an imminent machine failure.

Secondly, M&S provides a testbed for new event detection algorithms and faster than real-time studies. Event detection algorithms may be developed and implemented within an M&S framework more easily than within a real system, and simulation allows these implementations to be tested faster than in the real system. Of course, the M&S must be of high enough fidelity to be validated as being similar enough to the actual operating environment of the fielded event detection system.

Lastly, M&S may be used within an event detection algorithm implementation. Real-world data is often abstracted or simplified by using M&S. Andrade, Blunsden, and Fisher [2006], for instance, present an automatic technique for detecting abnormal events in crowds by abstracting the original data using M&S. Crowd behavior is typically difficult to predict or translate semantically. Moreover, it is difficult to track individuals in a crowd even when using state-of-the-art tracking algorithms. Thus, the authors characterize the crowd behavior by observing the crowd optical flow and use unsupervised feature extraction to encode normal crowd behavior. The unsupervised feature extraction applies spectral clustering to find the optimal number of models to represent normal crowd motion patterns. The motion models are Hidden Markov Models (HMMs) to cope with the variable number of motion samples that might be present within each observation window. The results of this technique clearly demonstrate its effectiveness in detecting crowd emergency situations.

Epilogue

This article defined event detection in the context of both natural and artificial systems. The primary challenges, methods, and applications of event detection were examined through research, examples, and literary references. Additionally, this article described the relationships between event detection and modeling and simulation. While much explanation and numerous examples were provided, this paper only serves to introduce each of these topics. The domain of applicability of event detection and its associated methods is expansive and continually growing.

As described, reliable event detection is a pervasive problem. Fortunately, there is no lack of researchers willing to study this problem, address its challenges, and devise innovative solutions. As such, event detection methods will continue to be an area of interest and much research now and into the future.

© 2009 Lockheed Martin MS2

References

Abadi D, Madden S, and Lindner W. 2005. "REED: Robust, Efficient Filtering and Event Detection in Sensor Networks." *Proceedings of the 31st Very Large Databases (VLDB) Conference*. Retrieved March 14, 2008, from <http://db.lcs.mit.edu/madden/html/reed_cr4.pdf>.

Andrade E, Blunsden S, Fisher R. 2006. "Modelling Crowd Scenes for Event Detection." *Proceedings of the International Conference on Pattern Recognition*, 1, 175–178. Retrieved March 13, 2009, from <http://homepages.inf.ed.ac.uk/rbf/PAPERS/andrade-crowd.pdf>.

Balazinska M. 2007. "Event Detection in Mobile Sensor Networks." *National Science Foundation (NSF) Workshop on Data Management for Mobile Sensor Networks (MobiSensors) 2007*. Retrieved July 25, 2008, from <http://mobisensors.cs.pitt.edu/files/papers/balazinska.pdf>.

Dash D, Margineantu D, and Wong WK. 2007. "Machine Learning Algorithms for Event Detection." A Special Issue of the *Machine Learning Journal*. Springer. Retrieved March 14, 2008, from <http://www.pittsburgh.intel-research.net/~dhdash/mlj_eventdetection.html>.

Favretto FO, Farias CRG, and Murta LO Jr. 2007. "A Decision Support System for Ischemic Event Detection." *Computers in Cardiology*, 34, 213–216. Retrieved March 14, 2008, from <http://www.cinc.org/Proceedings/2007/pdf/0213.pdf>.

Fienberg SE and Shmueli G. 2005. "Statistical Issues and Challenges associated with Rapid Detection of Bio-Terrorist Attacks." *Statistics in Medicine*, 24, 513-529. Retrieved September 12, 2008, from <http://www.niss.org/dgii/TR/FienbergShmueli-SIM-2005.pdf>.

Gupchup J, Burns R, Terzis A, and Szalay A. 2007. "Model-Based Event Detection in Wireless Sensor Networks." *Proceedings of the Workshop on Data Sharing and Interoperability on the World-Wide Sensor Web (DSI)*. Retrieved September 11, 2008, from <http://lifeunderyourfeet.org/en/literature/download/paper/LUYF_JHU_CR.pdf>.

Ihler A, Hutchins J, and Smyth P. 2006. "Adaptive Event Detection with Time-Varying Poisson Processes." *The Twelfth International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery)*. Retrieved February 28, 2008, from <http://www.ics.uci.edu/~ihler/papers/kdd06.pdf>.

International Council on Systems Engineering (INCOSE). 2006. *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities, Version 3*. San Diego, CA: International Council on Systems Engineering.

© 2009 Lockheed Martin MS2

Jasso H, Fountain T, Baru C, Hodgkiss W, Reich D, and Warner K. 2007. "Prediction of 9-1-1 Call Volumes for Emergency Event Detection." *The Proceedings of the 8th Annual International Digital Government Research Conference*. Retrieved March 17, 2008, from <http://scirad.sdsc.edu/datatech/data911/data911_files/dg.o_2007_paper.pdf>.

Kerman MC, Jiang W, Blumberg AF, and Buttrey SE. 2008. "A Comparison of Robust Metamodels for the Uncertainty Quantification of New York Harbor Oceanographic Data." *Journal of Operational Oceanography*, 1(2), 3-13.

Kerman MC, Jiang W, Blumberg AF, and Buttrey SE. 2009. "The Application of a Quantile Regression Metamodel for Salinity Event Detection Confirmation within New York Harbor Oceanographic Data." *Forthcoming*.

MSNBC News Service. 2008. "Crews fan out in Texas to search for Ike victims." September 14, 2008. Retrieved October 18, 2008, from <http://www.msnbc.msn.com/id/26637482/>.

NIST. 2008. *NIST/SEMATECH e-Handbook of Statistical Methods.* Retrieved October 18, 2008, from <http://www.itl.nist.gov/div898/handbook/index.htm>.

Sauvageon J, Agogino AM, Mehr AF, and Tumer IY. 2006. "Comparison of Event Detection Methods for Centralized Sensor Networks." *IEEE Sensors Applications Symposium 2006*.

Schwiderski-Grosche S. 2008. "Context-Dependent Event Detection in Sensor Networks." *The Second International Conference on Distributed Event-Based Systems (DEBS)*. Retrieved July 25, 2008, from <http://debs08.dis.uniroma1.it/pdf/fa-grosche-context.pdf>.

Sykes AO. 1993. "An Introduction to Regression Analysis." *Chicago Working Paper in Law & Economics*. University of Chicago Law School. Retrieved October 4, 2008, from <http://www.law.uchicago.edu/Lawecon/WkngPprs_01-25/20.Sykes.Regression.pdf>.

Tavakoli A, Zhang J, and Son S. 2005. "Group-Based Event Detection in Undersea Sensor Networks." *The Second International Workshop on Networked Sensing Systems*. Retrieved February 28, 2008, from <http://www.cs.virginia.edu/papers/GroupDetection_inss05.pdf>.

Tham CK. 2006. "Sensor-Grid Computing and SensorGrid Architecture for Event Detection, Classification and Decision-Making." *Sensor Network and Configuration: Fundamentals, Techniques, Platforms, and Experiments*. Springer-Verlag. Germany. Retrieved March 14, 2008, from <http://www.ece.nus.edu.sg/stfpage/eletck/sensorgrid/Springer%20CK%20Tham%20SensorGrid.pdf>.

Trafalis TB, Ince H, and Richman MB. 2003. "Tornado Detection with Support Vector Machines." *Lecture Notes in Computer Science*, 2660, 708.

© 2009 Lockheed Martin MS2

Welch G and Bishop G. 2006. “An Introduction to the Kalman Filter.” University of North Carolina at Chapel Hill, Chapel Hill, NC. TR 95-041. Retrieved October 4, 2008, from <http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf>.

Wikipedia. “Regression analysis.” Retrieved October 4, 2008, from <http://en.wikipedia.org/wiki/Regression_analysis>.

Zelen M. 2007. “The Early Detection of Disease – Statistical Challenges.” *Joint Statistical Meetings 2007*. Retrieved March 17, 2008, from <http://www.amstat.org/meetings/jsm/2007/webcasts/videos/JMS2007_FisherLecture/JSM2007_FisherLecture_files/fdeflt.htm>.

© 2009 Lockheed Martin MS2

Mitchell C. Kerman is a Principal Engineer at Lockheed Martin MS2 in Moorestown, New Jersey and a Ph.D. candidate in Systems Engineering at Stevens Institute of Technology in Hoboken, New Jersey. He has a B.S. in Computer Systems Engineering from Arizona State University and a M.S. in Operations Analysis from Naval Postgraduate School.

Wei Jiang is an Assistant Professor in the School of Systems & Enterprises at Stevens Institute of Technology. He has B.S. and M.S. degrees from Xi'an Jiaotong University and a Ph.D. from Hong Kong University of Science and Technology. He is currently the Chair-elect of the Data Mining section of the Institute for Operations Research and the Management Sciences (INFORMS).

Alan F. Blumberg is George Meade Bond Professor of Ocean Engineering at Stevens Institute of Technology. He leads the New York Harbor Observation and Prediction System which facilitates an assessment of ocean, weather, and environmental conditions throughout the New York Harbor region. He received a doctorate in Ocean Physics from The Johns Hopkins University and did post-doctoral work with Princeton University.

Samuel E. Buttrey is Associate Professor of Operations Research at Naval Postgraduate School in Monterey, California. He received a B.S. in Statistics from Princeton University and M.S. and Ph.D. degrees in Statistics from the University of California, Berkeley. His research interests include classification, clustering, large data sets, and statistical computing.

© 2009 Lockheed Martin MS2

Event Detection Challenges, Methods, and Applications in Natural and Artificial Systems

The 14th International Command and Control Research and Technology Symposium

Washington, DC

June 15-17, 2009

LOCKHEED MARTIN

Mitchell C. Kerman
Principal Engineer
Lockheed Martin MS2
609-326-5156
mitchell.c.kerman@lmco.com

© 2009 Lockheed Martin MS2

Agenda

- Introduction
  - Definition of a System
  - Classification of Systems
  - Definition of an Event
  - Event Detection
  - Sensor Employment
  - Static Threshold Event Detection
  - Research Goals
- Common Challenges in Event Detection
  - Situational Dependence
  - Criticality of Application
  - Numerous and Diverse Data Sources
  - Network Topology
  - Event Detection Algorithms
- Typical Event Detection Methods
  - Statistical Methods
  - Probabilistic Methods
  - Artificial Intelligence and Machine Learning Methods
  - Composite Methods
- Example Event Detection Applications
  - Network Monitoring
  - Health Monitoring and Management
  - Environmental Monitoring and Prediction
  - Safety and Security
  - Business Process Optimization
- Event Detection Modeling and Simulation
  - Requirements Development
  - Algorithm Testing
  - System Implementation
- Epilogue
- References

© 2009 Lockheed Martin MS2

INTRODUCTION

© 2009 Lockheed Martin MS2

Definition of a System

- System
  - "A combination of interacting elements organized to achieve one or more stated purposes" [INCOSE, 2006]
  - A collection of elements that, in combination, produce results generally not obtainable by the elements acting alone
    - Elements: Operators, hardware, software, firmware, information, policies, documents, techniques, facilities, services, and other support components
      - All items required to produce system-level results
    - System-level results: Qualities, properties, characteristics, functions, behaviors, and performance of the entire system

A system produces a desired behavior beyond the capacity of any individual system element or subgroup of system elements

© 2009 Lockheed Martin MS2

Classification of Systems

- Main Category
  - Natural
    - May not have an apparent objective
    - System inputs and outputs can be interpreted as serving a purpose
  - Artificial (or Man-made)
    - Designed for a specific purpose
    - Achieved through the delivery of outputs or services
- Subcategory
  - Observable
    - System inputs and outputs may be directly perceived in real-time
  - Non-observable
    - Either or both the system inputs and outputs may not be directly observed
- Method of Analysis
  - Qualitative
    - Delivery of the outputs
  - Quantitative
    - Measurement and analysis of specific system performance and effectiveness metrics derived from the system outputs

© 2009 Lockheed Martin MS2

Definition of an Event

- Event: A significant occurrence or large-scale activity that is unusual relative to normal patterns of behavior. May be associated with naturally occurring phenomena and manual system interactions.
  - Naturally occurring phenomena
    - e.g., Chemical and thermodynamic reactions and physical processes
  - Manual system interaction
    - e.g., An operator pushing a button
- An event results in the aberration of system parameters and output metrics
- Examples of events [Ihler, Hutchins, and Smyth, 2006]
  - A large meeting in an office building
  - A malicious attack on a Web server
  - A traffic accident on a freeway

Events are identified through a process known as "event detection"

© 2009 Lockheed Martin MS2

Event Detection

- **Observable systems**
  - **Direct observation of the system states**
  - **e.g., Looking outside to see if it is raining**
- **Non-observable systems**
  - **Sensors track the states of the parameters of interest**
  - **e.g., Using a thermometer to see if the outside temperature is below freezing**

© 2009 Lockheed Martin MS2

Sensor Employment

- Sensors
  - Organic (to the detection platform)
  - Local
  - Remote
  - Any combination of these
- Sensor outputs are inputs to event detection systems
- Regardless of the system, sensor-based event detection is among the most difficult and time-constrained of analysis problems
  - Requires excessive computational power
  - Consumes large amounts of storage space for voluminous data
- Example events detected using sensor-based event detection
  - A substantial change in sea level
  - An increase in background radiation level
  - The maneuver (or course change) of an anti-ship missile
  - An increase in pressure within a boiler (or heat exchanger)

© 2009 Lockheed Martin MS2

Static Threshold Event Detection

- Various methods of sensor-based event detection exist
- Static threshold event detection is one of the simplest and most common
  - e.g., Automobile fuel level sensor
- Simple method, but typically less reliable than more advanced techniques
  - e.g., What if the automobile fuel level sensor fails?
  - Many systems employ multiple (or redundant) sensors to overcome the reliability issues associated with a single sensor
    - Add complexity to the event detection problem since multiple inputs must be evaluated in order to determine whether or not an event is transpiring

© 2009 Lockheed Martin MS2

Research Goals

- Introduce the most common difficulties and challenges in event detection problems
- Describe the event detection methods most frequently employed
- Provide example event detection applications
- Explore the relationship between event detection and modeling and simulation

- This presentation incorporates the discoveries of and lessons learned by multiple researchers and authors over many combined years of experience in event detection theory and application
- This rather broad study has never been previously published within a single volume

© 2009 Lockheed Martin MS2

COMMON CHALLENGES IN EVENT DETECTION

© 2009 Lockheed Martin MS2

Situational Dependence

- Event detection problems are extremely situationally-dependent
- Several problems may be similar, but no two problems are ever exactly the same
  - Parameters, variables, and output metrics are selected based upon the specific event detection problem
  - Artifacts may or may not be applicable to other problems within the same domain, even for very closely-related problems

Approach in one domain may inspire alternative methods within other domains

© 2009 Lockheed Martin MS2

Criticality of Application

- Problems often address the requirements of a critical application
- e.g., Monitoring critical assets, measuring indicators of imminent catastrophic machine failures, detecting breaches within security perimeters, and observing human vital signs
- Require high precision and extreme timelines
  - High precision: A high true positive (i.e., correct detection) rate and a low false positive (i.e., incorrect detection) rate
  - Extreme timeline: A very short period of time in which the event detection method is able to correctly identify events
    - May range from less than a second to several minutes in duration (application dependent)

Event detection method must operate in real-time and fast enough to address the criticality of the application so that the detection report is not too time-late for an action or reaction to occur

© 2009 Lockheed Martin MS2

Numerous and Diverse Data Sources

- Any single event detection problem may consider a variety of diverse data sources with different data types and formats
  - Digital revolution exploded the number of data sources and amount of data readily available
  - Problem is compounded in assessing what data is actually relevant and approach must be capable of evaluating data from selected sources
    - Data must be aggregated, converted, or reformatted into a uniform structure that is independent of the data source
- Enormous volumes of data, often measuring in terabytes
  - Requires high-powered computing machinery and immense digital storage space
- Size of data set
  - Too little data can lead to missed detections or the development of an event detection solution which does not work in all cases
  - Too much data can lead to "analysis paralysis"
    - Detection problem is over-analyzed and never really solved
- Raw sensor data
  - Often plagued by inaccuracies and incompleteness
    - Inaccurate or missing position information
    - Delayed or out-of-order arrivals at receiving station
  - May exhibit cyclical, seasonal, and irregular trends
  - Often corrupted by a number of "burst" periods of atypical or unusual behavior

© 2009 Lockheed Martin MS2

Network Topology

- Network: A system containing a number of transmitting and receiving sensor stations, or nodes, that are connected through cables, wires, or wireless communications medium
- Network topology considers the locations and connectivity of these sensors in relation to the entire sensor network over time
  - In remote and mobile sensor networks, the network topology changes continuously due to sensor mobility and sensor lifetime
- Care and maintenance of the sensor network
  - Motes, or remote sensor nodes within a wireless sensor network, require maintenance and reseeding due to movement outside of the intended observed area, power consumption, sensor failures, and finite sensor lifetimes
- Network throughput and capacity
  - Aggregated data
    - Increases network throughput and reduces data processing times, but can significantly reduce the chance of detection since data from unaffected areas can mask the event signature
    - Detection system takes longer to notice the slight change in the aggregated data
  - Localized, sensor-level data
    - Improves detection sensitivity, but processing time for larger volumes of data can affect timeliness of detection
- Other considerations
  - Event persistence: The number of positive sensor detections required (from the same sensor) in order to report the occurrence of an event
  - Event lifetime: The length of an event as determined by the event persistence algorithm in signaling the start and end of the event
  - Context fluttering: An event indication is activated and deactivated in close succession due to inaccurate sensor readings or network delays
    - Sensor hunting: Activation and deactivation in close succession due to errors in a single sensor

© 2009 Lockheed Martin MS2

Event Detection Algorithms

- Three main requirements: Timeliness, a high true detection rate, and a low false alarm rate
- Timeliness
  - Implies immediate analysis of incoming data and immediate reporting of the results
  - Fast storage and analysis are critical
  - Detection algorithm must be efficient (i.e., fast and computationally cheap)
  - May give priority to some solution approaches over others
- Initialization, learning, and stabilization times
  - Time for the algorithm parameters to properly initialize, learn from the event-free environment, and then reach a stable state
  - Preliminary (learning) data for the algorithm must be known to be void of the events of interest
  - Detection system "learns" to detect the event based upon the event-free situation
  - Disadvantage
    - A "day zero" event: An event which is uncharacteristic of the normal events and has a never-seen-before signature
    - Detection algorithm has no means to detect a "day zero" event, as the algorithm is not actively "looking" for it
- "Roll-forward" approach
  - Each new data point is assessed for the indication of an event as it is added to the data set
  - Detection system should output an operational decision-making conclusion upon completion of the analysis
- Precision vs. Recall trade-off
  - Precision: The fraction of reported events that are actual (true) events
  - Recall: The fraction of all events that are reported correctly
  - In a pessimistic approach, the algorithm ignores large numbers of (potential) events due to the data uncertainty
    - Precision is high, but many actual events are missed, reducing the recall value
  - In the optimistic approach, events are reported even in the presence of uncertain data
    - Precision is lower since the errors in the input data result in false events, but the recall value is higher since fewer true events are missed
- Active adversary
  - Many event detection problems are exacerbated by the presence of an active adversary

© 2009 Lockheed Martin MS2

TYPICAL EVENT DETECTION METHODS

© 2009 Lockheed Martin MS2

Event Detection Methods

- No clear manner in which to characterize every event detection method
- Typical event detection methods may be classified into four rather broad categories
  - Statistical
  - Probabilistic
  - Artificial Intelligence and Machine Learning
  - Composite

© 2009 Lockheed Martin MS2

Statistical Methods (1 of 2)

- Static threshold method
  - Simplest and most computationally straight-forward
  - Detections are reported when the monitored parameter exceeds a predetermined threshold value
  - Detection condition persists as long as the parameter value exceeds the threshold set point
  - Threshold values may be determined based upon historical parameter values, analogy to similar sensors and systems, engineering estimates, or parametric analysis
- Regression
  - A data modeling and analysis technique in which the dependent variable is modeled as a function of independent variables, constant parameters, and an error term
    - Error term represents the variation in the dependent variable that cannot be explained by the model
  - Linear regression
    - Models the relationship between the dependent and independent variables as a straight line
  - Polynomial regression
    - Models the relationship between the dependent and independent variables as a polynomial
  - LOESS regression
    - Locally weighted regression
    - Fits a regression surface to data by multivariate smoothing
      - Simple models are fit to local subsets of data
  - Quantile regression
    - Estimates models for any of the conditional quantiles by minimizing sums of absolute residuals
    - Provides a more complete statistical analysis of the stochastic relationships among random variables

© 2009 Lockheed Martin MS2

Statistical Methods (2 of 2)

- Time series analysis
  - Time series: A sequence of successive data points typically separated by a uniform time interval
  - Three broad model classes
    - Autoregressive (AR)
    - Integrated (I)
    - Moving average (MA)
  - Composite models
    - Autoregressive moving average (ARMA)
    - Autoregressive integrated moving average (ARIMA)
- Kalman filter
  - An efficient recursive filter that estimates the state of a dynamic system from a series of incomplete and noisy measurements
- Model fitting interpolation
  - Interpolate values at intermediate points
  - e.g., use the bicubic technique to interpolate the value at a point as the weighted average of its nearest sixteen neighbor points
- Principal Component Analysis (PCA)
  - Also known as the Karhunen-Loève transform (KLT)
  - Uses singular value decomposition (SVD) to reduce high-dimensional datasets into datasets with lower dimensions that approximate the original data

© 2009 Lockheed Martin MS2

Probabilistic Methods

- Techniques in which the probability of event occurrence and other related probabilities and parameters are computed and assessed rather than computing and testing statistics from a sample data set
- Time-varying Poisson process model
  - Adaptively separates unusual event plumes from normal activity
  - Accounts for anomalous events
  - Outperforms the static threshold-based event detection technique
- Distributed Gaussian Method (DGM)
  - Generates Gaussian curves centered on each node
  - Curves are normalized and summed to reduce the geometric effect of node placement
  - Maximum value is then easily located
- SensorGrid [Tham, 2006]
  - An architecture for integrating sensor networks with grid computing
  - Grid computing involves groups of heterogeneous computational servers connected via high-speed network connections
  - Real-time information is mined, extracted, correlated, and processed to facilitate “on-the-fly” decisions and actions
  - Architecture relies upon distributed data fusion, event detection, and classification via probabilistic algorithms

© 2009 Lockheed Martin MS2

Artificial Intelligence and Machine Learning Methods

- Usually both computationally and informationally intensive
- Sensor sources are often sparsely distributed in time and space
  - Require advanced fusion algorithms to correlate the data from multiple sources
- Database operations
  - The most direct of these methods
  - Includes database queries and table joins
- Mote Fuzzy Validation and Fusion (Mote-FVF)
  - Developed for wireless sensors network
  - Can distinguish between sensor failures and abnormal environmental behaviors by using network redundancy to compensate for sensor reliability
  - Does not require or rely upon a mathematical model of the system
- Particle filtering
- Genetic algorithms
- Neural networks
- Intelligent agents

© 2009 Lockheed Martin MS2

Composite Methods

- Those methods that combine techniques within a category or from two or more of the categories
- Bayesian Gaussian Process (BGP) models
  - Combine probabilistic and machine learning methods
  - Powerful non-parametric learning methods based on simple probabilistic models

© 2009 Lockheed Martin MS2

EXAMPLE EVENT DETECTION APPLICATIONS

© 2009 Lockheed Martin MS2

Network Monitoring

- Monitoring Internet connections and conducting Web access logging
  - Frequency of visits to websites
  - General geographic locations of website visitors
  - Internet usage by employees
  - Security of online systems
    - Website intrusion detection
    - Failed account access logging
- Traffic monitoring
  - Determine whether or not an intersection requires a traffic signal

© 2009 Lockheed Martin MS2

Health Monitoring and Management

- Epidemic (or pandemic) detection and prevention
  - Center for Disease Control and Prevention (CDC) continuously monitors medical and public health information from physicians and hospitals across the country
  - Goals
    - Earliest possible detection of viruses and disease
    - Halt the spread by quarantining and treating the afflicted individuals
  - Afflictions of interest
    - Naturally occurring, such as the influenza virus
    - Bio-terrorist developed/released
- Early detection of disease within individual patients
  - Screening and monitoring programs
    - Diseases such as diabetes, hypertension, thyroid disease, tuberculosis, cancer, and coronary artery disease
    - Age to begin screening exams, the intervals between exams, and (possibly) the age to end screening exams
  - Diagnose and treat patients before they show any signs or symptoms (i.e., while in the pre-clinical state)
- Aerospace applications
  - Timely detection of local health anomalies has a great impact on the safety of the mission

© 2009 Lockheed Martin MS2

Environmental Monitoring and Prediction

- Early warnings of impending natural disasters
  - Tornadoes
  - Hurricanes
  - Tsunamis
  - Earthquakes
  - Floods
  - Volcanic eruptions
- Contamination of natural resources
  - Potable water is continuously monitored by water utilities for purity and potential contaminants
    - Causes
      - Natural
      - Man-made (e.g., terrorist)

© 2009 Lockheed Martin MS2

Safety and Security

- **Physical intrusion detection**
  - Home and corporate security alarm systems
- **Fire safety**
  - Fire, smoke, and carbon monoxide alarm systems
- **Homeland security**
  - Cargo security
    - Verify that the contents of cargo was not compromised during shipment
  - Threat detection and management
    - Detection, tracking, and interception of threat missiles is a quintessential military threat management example
    - Intrusion detection of enemy submarines within an operating area
- **Prediction of 9-1-1 call volumes**
  - Aids emergency service providers in service planning and recognition of anomalous calls

© 2009 Lockheed Martin MS2

Business Process Optimization

- Manufacturers rely heavily upon event detection methods
  - Reduce overall maintenance costs
    - Manufacturing and condition-based maintenance
      - Identify machines or processes that are in need of repair or adjustment
  - Ensure compliance with requirements
    - Business process compliance
      - Food and drug manufacturing
        - » Strict regulatory requirements obligate companies to certify that products do not exceed specific environmental parameters during processing

© 2009 Lockheed Martin MS2

EVENT DETECTION MODELING AND SIMULATION

© 2009 Lockheed Martin MS2

Relationship between Event Detection and Modeling and Simulation

- Intimate relationship and indivisible link between Event Detection and Modeling and Simulation (M&S)
  - Requirements Development
  - Algorithm Testing
  - System Implementation

© 2009 Lockheed Martin MS2

Requirements Development

- Use M&S at the forefront of the systems engineering process as a requirements development tool for an event detection system
- Requires a detailed study of the real-world system
  - Examine the parameters of interest
  - Understand the relationships between the system inputs and outputs
  - Gain deeper insight into the system interactions
- Through M&S, the systems engineer may determine what events can and need to be detected and what parameters must be monitored to detect these events
  - e.g., An engineer may determine that mechanical vibration and noise levels must be monitored as indications of an imminent machine failure

© 2009 Lockheed Martin MS2

Algorithm Testing

- M&S provides a test bed for new event detection algorithms and faster than real-time studies
  - Event detection algorithms are implemented within an M&S framework more easily than within a real system
  - Simulation allows the implementations to be tested faster than in the real system
- Caveat: M&S must be of high enough fidelity to be validated (as similar enough to the actual operating environment of the fielded event detection system)

© 2009 Lockheed Martin MS2

System Implementation

- M&S may be used within an event detection system implementation to abstract or simplify real-world data
- Andrade, Blunsden, and Fisher [2006] present an automatic technique for detecting abnormal events in crowds by abstracting the original data using M&S
  - Crowd behavior is typically difficult to predict or translate semantically
  - It is also difficult to track individuals in a crowd even when using state-of-the-art tracking algorithms
  - Characterize crowd behavior by observing the crowd optical flow and use unsupervised feature extraction to encode normal crowd behavior
  - Unsupervised feature extraction applies spectral clustering to find the optimal number of models to represent normal crowd motion patterns
  - Crowd motion models are Hidden Markov Models (HMMs) to cope with the variable number of motion samples that might be present within each observation window
  - Results of this technique clearly demonstrate its effectiveness in detecting crowd emergency situations

© 2009 Lockheed Martin MS2

EPILOGUE

© 2009 Lockheed Martin MS2

Summary

- This presentation merely scratched the surface of event detection challenges, methods, and applications
  - The domain of applicability of event detection and its associated methods is expansive and ever increasing
- Reliable event detection is a pervasive problem
  - Requires detailed problem analysis and innovative solutions to overcome a myriad of challenges
  - Fortunately, there is no lack of researchers willing to accept these challenges
- Event detection methods will continue to be an area of interest and much research now and into the future

© 2009 Lockheed Martin MS2

REFERENCES

© 2009 Lockheed Martin MS2

References

- Abadi D, Madden S, and Lindner W. 2005. "REED: Robust, Efficient Filtering and Event Detection in Sensor Networks." *Proceedings of the 31st Very Large Databases (VLDB) Conference*. Retrieved March 14, 2008, from <http://db.lcs.mit.edu/madden/html/reed_cr4.pdf>.
- Andrade E, Blunsden S, Fisher R. 2006. "Modelling Crowd Scenes for Event Detection." *Proceedings of the International Conference on Pattern Recognition*, 1, 175–178. Retrieved March 13, 2009, from <http://homepages.inf.ed.ac.uk/rbf/PAPERS/andrade-crowd.pdf>.
- Balazinska M. 2007. "Event Detection in Mobile Sensor Networks." *National Science Foundation (NSF) Workshop on Data Management for Mobile Sensor Networks (MobiSensors) 2007*. Retrieved July 25, 2008, from <http://mobisensors.cs.pitt.edu/files/papers/balazinska.pdf>.
- Dash D, Margineantu D, and Wong WK. 2007. "Machine Learning Algorithms for Event Detection." A Special Issue of the *Machine Learning Journal*. Springer. Retrieved March 14, 2008, from <http://www.pittsburgh.intel-research.net/~dhdash/mlj_eventdetection.html>.
- Favretto FO, Farias CRG, and Murta LO Jr. 2007. "A Decision Support System for Ischemic Event Detection." *Computers in Cardiology*, 34, 213–216. Retrieved March 14, 2008, from <http://www.cinc.org/Proceedings/2007/pdf/0213.pdf>.
- Fienberg SE and Shmueli G. 2005. "Statistical Issues and Challenges associated with Rapid Detection of Bio-Terrorist Attacks." *Statistics in Medicine*, 24, 513-529. Retrieved September 12, 2008, from <http://www.niss.org/dgii/TR/FienbergShmueli-SIM-2005.pdf>.
- Gupchup J, Burns R, Terzis A, and Szalay A. 2007. "Model-Based Event Detection in Wireless Sensor Networks." *Proceedings of the Workshop on Data Sharing and Interoperability on the World-Wide Sensor Web (DSI)*. Retrieved September 11, 2008, from <http://lifeunderyourfeet.org/en/literature/download/paper/LUYF_JHU_CR.pdf>.
- Ihler A, Hutchins J, and Smyth P. 2006. "Adaptive Event Detection with Time-Varying Poisson Processes." *The Twelfth International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery)*. Retrieved February 28, 2008, from <http://www.ics.uci.edu/~ihler/papers/kdd06.pdf>.
- International Council on Systems Engineering (INCOSE). 2006. *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities, Version 3*. San Diego, CA: International Council on Systems Engineering.
- Jasso H, Fountain T, Baru C, Hodgkiss W, Reich D, and Warner K. 2007. "Prediction of 9-1-1 Call Volumes for Emergency Event Detection." *The Proceedings of the 8th Annual International Digital Government Research Conference*. Retrieved March 17, 2008, from <http://scirad.sdsc.edu/datatech/data911/data911_files/dg.o_2007_paper.pdf>.
- Kerman MC, Jiang W, Blumberg AF, and Buttrey SE. 2008. "A Comparison of Robust Metamodels for the Uncertainty Quantification of New York Harbor Oceanographic Data." *Journal of Operational Oceanography*, 1(2), 3-13.
- Kerman MC, Jiang W, Blumberg AF, and Buttrey SE. 2009. "The Application of a Quantile Regression Metamodel for Salinity Event Detection Confirmation within New York Harbor Oceanographic Data." *Forthcoming*.
- MSNBC News Service. 2008. "Crews fan out in Texas to search for Ike victims." September 14, 2008. Retrieved October 18, 2008, from <http://www.msnbc.msn.com/id/26637482/>.
- NIST. 2008. *NIST/SEMATECH e-Handbook of Statistical Methods*. Retrieved October 18, 2008, from <http://www.itl.nist.gov/div898/handbook/index.htm>.
- Sauvageon J, Agogino AM, Mehr AF, and Tumer IY. 2006. "Comparison of Event Detection Methods for Centralized Sensor Networks." *IEEE Sensors Applications Symposium 2006*.
- Schwiderski-Grosche S. 2008. "Context-Dependent Event Detection in Sensor Networks." *The Second International Conference on Distributed Event-Based Systems (DEBS)*. Retrieved July 25, 2008, from <http://debs08.dis.uniroma1.it/pdf/fa-grosche-context.pdf>.
- Sykes AO. 1993. "An Introduction to Regression Analysis." *Chicago Working Paper in Law & Economics*. University of Chicago Law School. Retrieved October 4, 2008, from <http://www.law.uchicago.edu/Lawecon/WkngPprs_01-25/20.Sykes.Regression.pdf>.
- Tavakoli A, Zhang J, and Son S. 2005. "Group-Based Event Detection in Undersea Sensor Networks." *The Second International Workshop on Networked Sensing Systems*. Retrieved February 28, 2008, from <http://www.cs.virginia.edu/papers/GroupDetection_inss05.pdf>.
- Tham CK. 2006. "Sensor-Grid Computing and SensorGrid Architecture for Event Detection, Classification and Decision-Making." *Sensor Network and Configuration: Fundamentals, Techniques, Platforms, and Experiments*. Springer-Verlag. Germany. Retrieved March 14, 2008, from <http://www.ece.nus.edu.sg/stfpage/eletck/sensorgrid/Springer%20CK%20Tham%20SensorGrid.pdf>.
- Trafalis TB, Ince H, and Richman MB. 2003. "Tornado Detection with Support Vector Machines." *Lecture Notes in Computer Science*, 2660, 708.
- Welch G and Bishop G. 2006. "An Introduction to the Kalman Filter." University of North Carolina at Chapel Hill, Chapel Hill, NC. TR 95-041. Retrieved October 4, 2008, from <http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf>.
- Wikipedia. "Regression analysis." Retrieved October 4, 2008, from <http://en.wikipedia.org/wiki/Regression_analysis>.
- Zelen M. 2007. "The Early Detection of Disease – Statistical Challenges." *Joint Statistical Meetings 2007*. Retrieved March 17, 2008, from <http://www.amstat.org/meetings/jsm/2007/webcasts/videos/JMS2007_FisherLecture/JSM2007_FisherLecture_files/fdeflt.htm>.

© 2009 Lockheed Martin MS2